General Comments:

A. Overall, the memo contains a well-presented approach that is generally consistent with guidance developed by EPA Region 2 to support the Baseline Ecological Risk Assessment (BERA) for the 17-mile Lower Passaic River (LPR) Remedial Investigation (RI). Please note that the SQT analysis in the Draft Final 17-mile BERA is still being reviewed by EPA and the partner agencies, and it is anticipated that some significant changes will be necessary to finalize the Windward analysis. Agency review of the Draft Final LPR BERA (Windward, 2016) determined that the SQT focuses heavily on classification of impacted sediments on a sample-by-sample evaluation that treats site and reference data independently and reduces the power to detect potentially meaningful dose-response-like relationships. Tierra should instead combine site and reference data for comprehensive statistical evaluations that could reveal dose-response-like relationships.

B. EPA concluded that the reference envelope analysis reported in the Draft Final LPR BERA, which incorporated ad-hoc weights of evidence, arbitrary rankings, and many sequential tests to reduce the number of metrics, chemicals and locations, was not sufficiently robust to support remedial decision-making for the site. Rather the LPR Cooperating Parties Group (CPG) was directed to supplement the reference envelope-based analysis with a more descriptive, integrated multivariate approach. Rather than rely primarily on binary hypothesis testing, the overall emphasis of the analysis should be to clearly answer a small number of basic research questions[1] with the goal of enabling decision-makers to derive robust risk management decisions over broad spatial scales. This direction is also applicable to the Newark Bay risk assessments and in addition, sufficient graphical presentation (including estimation statistics and confidence intervals) of the biological response datasets should be provided to facilitate understanding of potential differences between reference and site distributions as well as potential spatial differences within the Newark Bay Study Area (NBSA).

C. The limitations of the proposed data reduction and scoring procedures should be clearly stated and explicitly considered when interpreting the results. Specific areas of potential concern include:

- underappreciation of the impact of spatial heterogeneity of stressors in sediment;
- reliance on decisions based on unreplicated information due to individual sample-by-sample decision processes;
- understatement of impacted sites due to the requirements for biological metrics to be much lower than average conditions at the reference site (i.e., lower than the 5th percentile); and,
- over-reliance on yes-no tests of hypotheses rather than development of statements describing the magnitude of impact, including measures of uncertainty.

In general, Tierra should not proceed with the NBSA SQT analysis until the Windward assessment for the LPR 17-mile RI has been finalized and approved by EPA. This may entail some changes to the proposed methodology. Following this review and an in-depth evaluation of the SQT dataset, a conference call between Tierra and EPA may be warranted to discuss specific analyses and reach consensus on outstanding questions.

[1] Some relevant research questions are provided at the end of the memorandum.

Specific Comments:

1. Section 1 - "Background", p. 2/18, last paragraph. Tierra is reminded that additional data collection activities may be necessary to support the BERA, and the SQT specifically, in the event that it is determined that the datasets lack adequate statistical power to meet established data use objectives. There were several discussions between EPA and Tierra regarding the comparatively low data density (sediment and SQT stations) relative to other contaminated sediment sites (including the LPR) during project scoping. As part of initial data evaluation activities, Tierra should review the statistical analysis used to derive the number (i.e., 30) of SQT stations and discuss the implications of any identified incorrect assumptions with USEPA.

2. Section 2 – "Objectives and Approach", General. Specific research questions (examples provided below, following comments) should be defined at the beginning of this section with details describing how the proposed analyses will answer the research questions.

3. Section 2 – "Objectives and Approach", p. 3/18, first paragraph after bullet items. Assigning equal weighting to both the sediment and porewater components of the chemistry LOE is consistent with EPA (2015) guidance; however, if the latter data exhibit stronger correlations with the biological response metrics than does the sediment chemistry, alternate weighting schemes should be discussed with EPA. It is noted that the scoring of sediment chemistry added little heuristic value to the Windward analysis as all site stations were similarly scored (i.e., a value of 1). Consequently, use of alternative sediment benchmarks that are better correlated with the biological response metrics, as suggested in the proposed approach, may be appropriate. Strong correlations between bulk sediment chemistry and biological metrics are unlikely whereas this may be the case for porewater chemistry. If so, Tierra should discuss alternative weighting schemes with EPA as either the primary analysis or as part of the uncertainty assessment.

4. Section 3.1 – "Chemistry Data", p. 4/18, first paragraph. Consistent with the Phase II Data Evaluation Report (DEAR), both raw and carbon-normalized sediment chemistry and raw and lipid-normalized tissue chemistry datasets should be presented.

5. Section 4 – "Sediment Quality Triad Evaluation", p. 5/18. Weisberg et al., (1998) discusses the atypical (relative to the other BIC metrics) nature of the abundance metric (both high and low values may be indicative of environmental stress) and provides a suggested alternative approach for scoring this metric. Consistent with Windward (2016), an upper envelope threshold based on the 95th percentile should be developed for this metric.

6. Section 4.1 – "Benthic Invertebrate Community", p. 6/18, top of page. Windward (2016) includes 31 (not 25) Jamaica Bay sampling locations in the urban reference dataset but did not include 4 locations sampled in 2003 that meet the combined chemistry and toxicity screening criteria. The complete set (n=35) of acceptable sampling locations should be used to develop the reference envelopes for NBSA biological response metrics.

7. Section 4.1 – "Benthic Invertebrate Community", p. 6/18, top of page. Justification for use of Jamaica Bay as a reference site should be clearly documented. The impact of using a reference dataset with *Ampelisca* to conduct the NBSA SQT, where *Leptocheirus* was the test species, should be evaluated in the uncertainty assessment. In addition to a review of comparative

toxicity data presented in the literature, an evaluation of laboratory toxicity data for these two species for samples collected in Jamaica Bay (REMAP and Newtown Creek reference locations including Head of Bay) may be useful.

8. Section 4.1 – "Benthic Invertebrate Community", p. 6/18, top of page. Additional details should be provided to describe how the SQT analysis will address the potential confounding issue of excessive nutrient loading to the Jamaica Bay reference area. In addition, this should be included as a topic of discussion in the uncertainty assessment.

9. Section 4.2 – "Sediment Toxicity", p. 7/18. With respect to defining control acceptability criteria, 90th percentile minimum significant differences (MSD) thresholds should be calculated consistent with Windward (2016) and used to identify stations with significantly different sample responses.

10. Section 4.2 – "Sediment Toxicity", p. 6-7/18. The referenced Eickhoff et al. (2014) presentation discusses options for reducing variability in the reproductive endpoint dataset (i.e., using a subset of results for samples that meet a minimum growth threshold value). Such an approach should be considered as an alternative or in addition to that proposed in the first complete paragraph on page 7/18.

11. Section 4.3 – "Sediment Chemistry". See above Comment No. 3 related to assigning equal weights to the sediment and porewater components of the chemistry LOE.

12. Section 4.3.1 – "Whole Sediment Chemistry", p 7/18. As indicated in the draft memorandum, reliability concerns related to use of T20/T50 values as observed in Windward (2016) may require that other sediment benchmark approaches be selected to score the sediment chemistry LOE. Compared to ERL/ERMs, T20 and T50 values appear to be conservatively low and may result in a high percentage of false positive determinations.

13. Section 4.3.1 – "Whole Sediment Chemistry", p 7/18. To support the scoring of the chemistry triad leg, sediment screening criteria should be identified for other contaminants that are found to be prominent in the sediment chemistry results but lack T20/T50/T80 values. Selected values should approximate the same level of conservatism to the extent possible. For TCDD, the apparent effects threshold (AET) – based marine sediment criterion of $3.6E10^{-6}$ mg/kg (available at https://repository.library.noaa.gov/view/noaa/9327).

14. Section 4.3.2 – "Porewater Chemistry", p. 8/18. A sum of toxic units approach should be applied for the SEM metals. In addition to the porewater assessment, AVS and SEM should also be considered for SEM. Pursuant to an EqP-based AVS and SEM approach for metals (USEPA, 2005b), it is sufficient to conclude that toxicity is not expected if the excess SEM criterion is satisfied for the SEM metals, even if the sum of toxic units for the dissolved metals exceeds unity.

15. Section 4.3.2 – "Porewater Chemistry", p. 8/18. For scoring the chemistry leg of the triad, alternative (i.e., acute and/or freshwater-based) water criteria should be identified for contaminants lacking a saline NJSWQS or chronic AWQC. For instance, for TCDD, since a marine criterion is not available, the freshwater aquatic chronic criterion of $3E10^{-9}$ ug/l (available at http://www.nj.gov/dep/srp/guidance/ecoscreening/) should be used. Please revise this section to also indicate how chemicals lacking aquatic threshold values will be accommodated in the

chemistry scoring process.

16. Section 5.2.3 – "Community Ordination", p.10/18. To supplement the proposed approach, Tierra should include use of other ordination techniques such as redundancy analysis (RDA) or canonical correspondence analysis (CCA) to directly evaluate the influence of physicochemical data (e.g., grainsize, depth, DO) on the distribution of benthic species in NBSA.

17. Table 2. The cited table from Windward (2016) contains multiple errors (e.g., sample size, reference envelope values). In addition, the urban reference dataset for Jamaica Bay is not complete. EPA is aware of 35 locations that meet the combined toxicity and chemistry screening criteria defined in EPA (2015) guidance. Tierra should ensure that the urban reference envelope is consistent with the finalized Windward SQT analysis and be based on a complete set of acceptable reference stations.

18. Figure 1. See specific Comments Nos. 13 and 15 regarding handling of chemicals lacking benchmarks. Terms like "highly correlated" should be defined. In addition, the individual elimination steps (including benchmark comparisons and correlations factors) appear reasonable; however, conducting the multivariate analyses (i.e., PCA/GLM) on a subset of data could result in meaningful relationships between variables being missed. Tierra also evaluate the full dataset (including both NBSA and Jamaica Bay) for selected endpoints and metrics as part of the statistical analysis to assess overall structural stability of the dataset.

<u>Example Research Questions</u>:

The following are examples of some recommended questions that should be answered so that risk managers can develop a full understanding of site conditions, the spatial scales at which toxic exposures may be operating, and the limitations of scale at which the existing data can resolve toxic conditions. Answers to these basic environmental questions should provide context for interpreting the proposed station classification scheme.

1. Do benthic invertebrate survival and growth rates at Newark Bay differ from those observed at other regionally available reference sites?

    a. What is the magnitude of the difference and how precisely can this difference be estimated (i.e., report mean differences in survival between the site and other reference areas, including 95% confidence limits on these differences.)

2. Do benthic invertebrate indices in geographical areas within the site vary and are indices in any areas similar to those at the reference area?

    a. Are growth and survival in the north end of the bay similar to that in the south end?
    b. Are growth and survival in intertidal flats similar to other hydrodynamic regimes?
    c. *Other similar questions are likely to be of interest.*

3. Are observed differences associated with chemistry or physical conditions (e.g., grain size and organic carbon) or both or neither?